DOCUMENT RESUME

ED 178 602 TM 009 821

AUTHOR Ruck, Hendrick W.
TITLE The Collection and Prediction of Training Emphasis Ratings for Curriculum Development.
PUB DATE [78]
NOTE 15p.; Paper presented at the annual meeting of the Military Testing Association (20th, Oklahoma City, OK, October 30-November 3, 1978); Appendix, pages 253-257, was removed prior to filming because print was not legible

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Curriculum; Educational Policy; Factor Analysis; *Job Analysis; *Job Training; Mathematical Models; *Military Training; *Needs Assessment; Policy Formation; Postsecondary Education; Predictive Validity; Rating Scales; Reliability; Specialization; *Training Objectives
IDENTIFIERS Air Force; Interrater Reliability

ABSTRACT

The generalizability of earlier findings on training objectives for military specialties was studied. Thirteen tasks or jobs, involving mechanical, electrical, general, or administrative aptitude, were rated by senior noncommissioned officers on: recommended field training emphasis; present school emphasis; consequences of inadequate job performance; delay tolerance; and task difficulty. Occupational survey data--concerning percent of members performing each task, percent of time spent on task, learning difficulty, and average grade of task performers--were also made available. Policy equations were used as models to "capture" the rankings of the judges' recommended training emphasis, and to validate the training emphasis scale. Factor and job analysis indicated that the two policies used yielded reliabilities of .72 and .64. Interrater agreement was also found to be related to predictive efficiency of the equations. Only a moderate number of raters (about 40) were needed to provide stable data, and the performance criterion used in this study was found to be reliable and valid. Further research is recommended for those complex specialties which had low predictive efficiency and poor interrater agreement. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178602

The Collection and Prediction of Training Emphasis Ratings for Curriculum Development

by

Hendrick W. Ruck
Nancy A. Thompson
David C. Thomson

Air Force Human Resources Laboratory
Brooks AFB, Texas

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The opinions and conclusions expressed in this paper are those of the authors and are not necessarily those of the United States Air Force.

One of the most difficult questions that arises in occupational curriculum design is, "What should the training content be?" This question, in the business of Air Force vocational training, could be further reduced to the fundamental questions of "Which occupational tasks should be included in the curriculum?" and "How do those tasks translate into specific skills and knowledge?" The purpose of this paper is to address the first of the fundamental questions; i.e., the selection of tasks for training.

The Air Force Human Resources Laboratory has been conducting extensive research in the training requirements area. The initial concepts and theory guiding the research were first proposed by Christal (1970), who suggested that boards of expert judges could be used to study information about tasks that are hypothesized to be related to the training decision. The experts could then evaluate those tasks in terms of the appropriateness for inclusion in curricula. He further suggested that the mathematical technique of policy capturing be applied to the judges' decisions so that the policy of the judges could be applied to additional tasks. This approach would reduce the necessity of expert judgment in task selection for each task and would assure more consistent decisions since the mathematical model of the experts' decisions could be used instead of additional judgments by the experts.

These initial suggestions have been studied in a stream of research on task training factors. Mial and Christal (1974) developed a number of task training factors and were able to predict judges' mean rank ordering of tasks for priority in training using a four-factor regression equation ($R=.88$; $P<.001$). Their research was conducted using the Medical Service specialty. Mead (1975) has presented additional evidence as to the utility of the policy-capturing approach. He performed a similar study to that of Mial and Christal using a different specialty (Law Enforcement) and met with similar success in predicting training priorities. Mead successfully used both mean rankings of training priority and mean ratings of training priority in his research. These studies suggested that a promising link between Instructional System

TM009821

Development (ISD) theory, occupational survey data, and curriculum design could be further developed.

Stacy, Thompson, and Thomson (1977) presented a paper last year at the Military Testing Association Conference outlining preliminary results of task factor data collection, training emphasis prediction, and task-anchored scaling. Stacy found that the task training factors could be collected reliably using standard occupational survey techniques. He also reported success in using the policy-capturing approach for a number of specialties. The present paper will discuss the results of policy-capturing research on 13 Air Force specialties, the similarities and differences in policies for different specialties, and the implications of the research for Instructional System Development (ISD). A separate paper is being presented at this conference by Squadron Leader David C. Thomson (Thomson and Goody, 1978) documenting the results of the task anchored scaling research.

Method

Research conducted prior to Stacy, et al (1977) focused on two specialties. This study was designed to test the generalizability of earlier findings. Therefore, 13 of the 14 specialties studied by Stacy, et al, were selected for this study (Table 1). The specialties were selected so that occupational survey data and job inventories were current, initial skill courses were mandatory for entry into the specialties, and all four aptitude areas used in Air Force job placement (Mechanical, Electrical, General, and Administrative) were represented. As a result of the operational occupational surveys conducted by the AF Occupational Measurement Center, data were available on percent members performing each task, an index of percent time spent on each task, the learning difficulty of each task, and the average grade of members performing each task. Additional data that were collected for the study included: (a) field recommended training emphasis for each task, (b) present school emphasis for each task, (c) probable consequences of inadequate performance for each task, and (d) delay tolerance for each task. The learning difficulty task factor used in this study was collected using a nine-point relative scale. However, the other two factors (consequences of inadequate performance, task delay tolerance) were collected using nine-point scales (Stacy, Thompson & Thomson; 1977) that were verbally anchored and did not require relative task comparisons. These factors have been described previously (Stacy, et al, 1977); however, the training emphasis scale will be described again in this paper because of its importance.

The field recommended training emphasis scale was developed as the criterion. It was expected to yield equivalent information to the mean rank orderings of training priority as used by Mial and Christal (1974), since Mead (1975) demonstrated the equivalency between rankings and ratings for training priority. The field recommended training emphasis scale is a nine-point scale ranging from "Extremely Little" to "Extremely Heavy." Senior NCOs serving in operational units in each specialty are

Table 1
AFSC Aptitude Areas and Raters

AFSC	Title	Aptitude Area	Number of Respondents/Raters: Members Total	Training Emphasis	Consequences	Delay	Difficulty
293X3	Radio Operator	A	1468	224	45	50	78
304X0	Radio Relay Equipment	E	1573	215	35	50	89
304X4	Ground Radio Communication Equipment	E	2351	335	60	58	122
328X3	Electronic Warfare	E	1223	306	46	47	43
472X2	General Purpose Vehicle Mechanic	M	3338	291	33	34	127
552X5	Plumbing Specialist	M	964	143	82	62	116
651X0	Procurement Specialist	A	979	320	61	63	101
672X1	General Accounting	A	596	85	55	55	86
672X2	Disbursement Accounting Specialist	A	1352	149	65	65	86
902X0	Medical Services	G	2198	380	93	95	58
906X0	Medical Administration	G	2356	300	105	104	78
911X0	Physiological Training	G	408	79	30	30	86
981X0	Dental Specialist	G	1856	89	65	47	45
Total			20662	4220	1096	1098	1451

asked to (a) check each task for which formal training (school or on-the-job training (OJT)) is recommended for first-term airmen, and (b) rate each of the tasks that were checked using the nine-point scale. The training emphasis scale is normally treated in data reduction and analysis as a 10-point scale since the absence of a check mark is treated as zero. This differs from other ISD task factors used in the Air Force occupational survey program since every task is normally considered to possess some amount of each ISD factor. That is, for example, no task would be expected to have zero learning difficulty. Similarly, no task would have zero consequences of inadequate performance or delay tolerance. Tasks could, however, have zero field recommended training emphasis.

The field recommended training emphasis scale has been intensively researched. It has been collected in the research mode for 19 specialties and in the operational mode for an additional 21 specialties. Table 2 lists the AF specialties and associated interrater agreement data for the field recommended training emphasis data collected to date. The median interrater agreement coefficient is .95. Analyses of rater agreement data suggest that a minimum of 40 raters should be used to provide reliable results for the recommended training emphasis scale.

The validation of field recommended training emphasis was performed using policy capturing (Christal, 1968). Policy capturing requires that a multiple regression model be developed in an attempt to "capture" the policy of the judges in their ratings or rankings. Basically, it is the development of explanatory and predictive regression models. The policy model that was developed to predict field recommended training emphasis included three task factors and three related job factors, together with squares of the factors. The task factors in the model were learning difficulty, probable consequences of inadequate performance, and task delay tolerance. The job-related factors were percent members performing in the first assignment, an index of percent time spent by members in their first assignment, and the average weighted grade of members performing. Since each factor was squared to address expected curvilinear relationships, a twelve variable regression equation was generated for each Air Force specialty.

The ISD literature has often been interpreted as suggesting that there is one correct way to combine task and job factor information in order to derive training requirements. This hypothesis was tested by analyzing the regression equations (Ward, 1963; Gott, 1978) for each specialty to determine whether, in fact, different policies as expressed in the policy equation exist across specialties, or, as one might expect, there is one universal equation (or combination rule). The analyses required to test the hypothesis can be conducted using a hierarchical grouping algorithm which tests similar regression equations for homogeneity of weights.

Table 2
Descriptive Statistics of Training Emphasis Ratings

AFSC	Title	Avg Mean	SD	R_{kk}*	No. of Raters
111X0	Defensive Aerial Gunner	3.49	1.94	.94	43
293X3	Ground Radio Operator	2.26	1.44	.92	189
303XX	Aircraft Control and Warning Radar	3.08	1.76	.96	52
304X0	Radio Relay Equipment	2.38	1.50	.94	199
304X4	Ground Radio Communication Equipment	1.82	1.15	.90	315
307X0	Telecommunication Systems Control	2.87	2.11	.96	75
321XX	Defensive Fire Control Systems	2.48	2.09	.97	50
328X3	Electronic Warfare Systems	1.02	1.25	.94	248
341XX	Training Devices	1.38	1.15	.89	46
361X0	Outside Wire and Antenna Maintenance	3.53	1.67	.94	40
423X1	Aircraft Environmental Systems	2.63	1.34	.91	137
423X4	Aircraft Pneudraulic Systems	2.89	1.55	.93	282
427X2	Nondestruction Inspection	3.89	2.29	.98	178
427X5	Airframe Repair	3.48	2.16	.97	267
443X0	Missile Maintenance (LGM 25-Titan)	3.91	2.30	.97	53
462X0	Aircraft Armament Systems	2.72	1.81	.96	186
463X0	Nuclear Weapons	2.05	2.29	.90	40
472X2	General Purpose Vehicle Maintenance	2.39	2.30	.98	243
472XX	Vehicle Maintenance	3.42	1.75	.95	49
542X2	Electronic Power Production	3.85	1.54	.93	40
552X5	Plumbing	3.32	1.62	.95	125
555X0	Programs and Work Control	2.61	1.48	.92	54
571X0	Fire Protection	3.55	1.87	.95	51
601X4	Packaging	1.42	1.72	.97	17
602XX	Passenger and Freight	2.75	1.58	.92	26
622X1	Diet Therapy	3.68	1.82	.95	47
631X0	Fuel	3.22	1.92	.95	277
645X0	Inventory Management	1.77	1.55	.92	12
645X1	Materiel	1.27	1.37	.92	15

Table 2 (Continued)
Descriptive Statistics of Training Emphasis Ratings

AFSC	Title	Avg Mean	SD	R_{kk}*	No. of Raters
645X2	Supply	1.31	2.25	.99	5
651X0	Procurement	2.85	1.63	.94	295
672X1	General Accounting	1.29	1.41	.95	73
672X2	Disbursement Accounting	1.21	1.37	.95	131
902X0	Medical Service	3.44	1.72	.93	302
904X0	Medical Laboratory	2.91	1.83	.95	46
906X0	Medical Administrative	1.71	1.19	.91	270
907X0	Environmental Health	3.25	1.82	.95	64
911X0	Aerospace Physiology	2.59	2.00	.96	68
981X0	Dental	2.33	2.06	.97	85
982X0	Dental Laboratory	2.84	2.02	.97	23

*Rater agreement indices for a sample of 40 raters as estimated by the Spearman Brown formula.

Analysis of Policy Equations

The results of the grouping analysis of the 13 policy equations are highlighted in Table 3. Notice that, if the regression equation derived for each of the 13 specialties was used to predict for that specialty, the overall predictive efficiency would be quite high (.86). On the other hand, predictive efficiency using a single averaged equation for each of the specialties would result in unacceptably low predictive efficiency of .56. As a result of the analysis, a compromise solution appears to be one which uses one equation (Policy A) for eight specialties, and a second equation (Policy B) for the remaining five specialties. The equation used in Policy A yields an R-squared of .72; however, the Policy B equation has an R-squared of .64. This suggests that the specialties in Policy B are not as predictable (using the ISD factors) as those in Policy A.

Table 3
Grouping of Training Priority Policy Equations

	Number of Equations	Overall Predictive Efficiency (R^2)
Maximum	13	.86
Optimal	2	.73
Minimum	1	.56

Additional analyses of the differences between the two policies were performed in an attempt to isolate characteristics of specialties in each policy. The specialties in Policy B differed from those in Policy A in that Policy B specialties were measured with job inventories that had significantly more tasks than in A ($\bar{X}_A = 425$, $\bar{X}_B = 951$; $t = 3.81$, $df = 11$, $p<.01$) and Policy B specialties included significantly more job types than Policy A ($\bar{X}_A = 27.1$, $\bar{X}_B = 47.6$, $t = -2.69$, $df = 11$, $p<.05$). It is important to note here that no significant relationship was found between number of tasks in a job inventory and number of job types identified in an occupational analysis ($r = .29$, ns).

Analysis of Interrater Agreement

Although the complex relationships among recommended training emphasis and the ISD factors will not be more fully developed in this presentation, one may conclude that there is no single method of combining ISD factor data to arrive at training emphasis for all specialties. This conclusion applies if Air Force specialties are considered the unit of analysis; however, the conclusion has not been tested, and may not hold up if the unit of analysis is changed to job groups within specialties rather than specialties. Nevertheless, the finding is

significant, since most technical training in the Air Force is developed for specialties and not for job groups.

After determining that there were at least two different policies that could be used to predict training emphasis, and that the policies differed in predictive efficiency and type of specialty, further analyses of interrater agreement were conducted. Although no difference in interrater agreement was found among the specialties in each policy group, interrater agreement was found to be moderately correlated with predictive efficiency. That is, the correlation between R-squared for each specialty and interrater agreement (R_{11}) on training emphasis is significant ($r = .61$, $p<.05$).

Table 2 displays the interrater agreement values adjusted for 40 raters. Using a conservative cutoff of .91 for acceptable interrater agreement, one can see that five (or 12.5 percent) of the specialties do not meet the cutoff. This analysis leads to the conclusion that the recommended training emphasis data can be reliably collected in at least 80 percent of the Air Force specialties.

Complex Specialties

The training emphasis research has resulted in a criterion that may be collected and used in decision making in a large number of specialties. Two types of problem specialties have been identified in the research. First, there are specialties with low interrater agreement. Second, there are specialties for which predictability of recommended training emphasis is not as high as is necessary for practical prediction. These complex specialties are being investigated further. The new research will attempt to (a) determine whether additional factors may be useful for predicting recommended training emphasis in complex specialties through the development of additional task factor scales, (b) examine the complex specialties for common characteristics, (c) determine optimal data displays for complex specialties for training decision makers, (d) determine which specialty characteristics are associated with low interrater agreement and poor predictability.

Discussion

The task training factor research stream has produced significant results. First, ISD task and job factors have been identified and scales developed to measure them. Second, the field recommended training emphasis scale has been developed as a criterion. The training emphasis scale has been shown to be reliable, through interrater agreement analyses in 40 specialties, and valid, through policy capturing and policy grouping in 13 specialties. Third, no single way of combining ISD factors for training decisions was found to be appropriate for all specialties.

The initial objective of the research was to discover combination rules that may be applied to ISD factor data for selecting tasks for

training. This has been done. However, the rules differ for different groups of specialties. The criterion used in the research has been found to be both reliable and valid. Furthermore, only a moderate number of raters (about 40) is required to provide stable data. These findings have led to the unexpected conclusion that the criterion should be collected and not predicted. It is important to note that recommended training emphasis ratings, although useful for most Air Force specialties, will not always be immediately usable. In particular, complex specialties appear to require additional study to enhance understanding of low predictive efficiency and poor interrater agreement.

As a result of this research, it has been recommended that supervisory ratings of formal training emphasis be collected routinely in the Air Force Occupational Survey Program. Further, it has been recommended that routine collection of the task factors Consequences of Inadequate Performance and Task Delay Tolerance be discontinued since recommended training emphasis ratings include consideration of those factors. In cases where more than one Air Force specialty would be included in a single job inventory, it is recommended that separate ratings be collected for each specialty and that those ratings be analyzed and presented for each of the specialties. Finally it has been recommended that the training emphasis data be presented using new modularized CODAP (Thew & Weissmueller, 1978) programs that allow a merging of training documentation, such as Specialty Training Standards (STS) or Plans of Instruction (POI) with job inventory tasks. This merging of job inventory tasks and training documents provides a simple and reliable method of displaying occupational survey data within the context that training personnel are most familiar. The Appendix displays an example of the output.

The research leading to the conclusions and recommendations has been difficult and complex. However, the available technology for using occupational survey data for training decisions has been considerably expanded and implementation of the results would provide a much stronger basis for making training decisions than is currently available.

References

Christal, R.E. Implications of Air Force occupational research for curriculum design. In B. B. Smith & J. Moss, Jr. (Eds.), Report of a seminar: Process and Techniques of vocational curriculum development. Minnesota Research Coordination Unit for Vocational Education, University of Minnesota, Minneapolis, MN, April 1970, 27-61.

Christal, R.E. JAN: A technique for analyzing group judgment. Journal of Experimental Education, 1968, 36(4), 24-27.

Gott, C.D. HIER-GRP: a computer program for the hierarchical grouping of regression equations. AFHRL-TR-78-14. Brooks Air Force Base, TX: Computational Sciences Division, Air Force Human Resources Laboratory, June 1978.

Mead, D.F. Determining training priorities for job tasks. Paper presented to 17th Annual Conference of the Military Testing Association, U.S. Army, Indianapolis, IN, 16-19 September 1975.

Mial, R.P. & Christal, R.E. The determination of training priority for vocational tasks. Proceedings, Psychology in the Air Force Symposium. USAF Academy, April 1974, 29-33.

Stacy, W.J., Thompson, N.A. & Thomson, D.C. Occupational task factors for instructional systems development. Paper presented to 19th Annual Conference of the Military Testing Association, USAF, San Antonio, TX, 17-21 October 1977.

Thew, M.C. & Weissmuller, J.J. CODAP: a modular approach to occupational analysis. Paper presented to 20th Annual Conference of the Military Testing Association, U.S. Coast Guard, Oklahoma City, OK, 30 October - 3 November 1978.

Thompson, D.C. & Goody, K. Benchmark scales for collecting task training factor data. Paper presented to 20th Annual Conference of the Military Testing Association, U.S. Coast Guard, Oklahoma City, OK, 30 October-3 November 1978.

Ward, J.H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 1963, 58, 236-244.

* APPENDIX

1. This appendix contains sample computer output that is being recommended for use by training managers and curriculum developers. The printout is in two parts; the first part (pp 12-14) is an executive summary, and the second part (pp 15-16) provides detailed Occupational Survey (OS) data that have been matched with Specialty Training Standard (STS) items.

2. Four columns are on each of the subsequent printouts. Column 1 displays the number of OS tasks that have been matched to each of the STS items. The data displayed in column 2 are the recommended training emphasis ratings collected from AFS 293X3 (Ground Radio Operator) field supervisors. Column 3 includes the percent of job incumbents with 2-24 months military service in the 293X3 career ladder who perform the tasks. Finally, Column 4 shows the learning difficulty for each task.

3. The executive summary has been designed to aggregate OS task data to STS item level. Pages 12-14 display STS items in original STS sequence. Mean values for each of the three OS task factors for the tasks that were matched to the STS item are displayed in columns 2-4. Note that in the case where no tasks have been matched with an STS item, the values in the adjacent columns are zero. Also, it is possible to show the same STS items and OS data with the STS items arranged in descending order on field supervisors' recommended training emphasis. This display gives a powerful overview of the data and is expected to be quite useful to training managers.

4. The remaining two pages (pages 15-16) provide samples of detailed OS data in the STS framework. STS items are printed between dashed lines and OS tasks (and associated data) are shown immediately below the items. Tasks are listed in order of training priority for introductory airmen within each STS category. The matching between STS and OS tasks was performed and reviewed by course personnel. Note that tasks may be mapped into as many STS items as required and that both STS items and OS tasks may have no counterparts.

* The Appendix (pp. 253-7) was removed prior to submission to the ERIC Document Reproduction Service, because the print was not legible.